24

Claims

1. Method for determining and outputting a similarity measure between two data strings each data string comprising data entities, comprising:
-   receiving a first data string,
-   receiving a second data string,
    characterized by
-   determining consecutively following data entities in said first data string,
-   determining the relative positions of said consecutively following data entities in said first data string
-   determining similar data entities with the same order in said second data string,
-   determining the relative positions of said determined data entities in said second data string,
-   determining a matching measure by determining how far the relative positions of data entities in said second data string match with the relative positions of consecutively following data entities in said first data string, and
-   outputting a similarity measure which corresponds to the matching measure of at least one comparison result.

2. Method according to claim 1, wherein pairs of consecutively following data entities are determined in said first data string.

3. Method according to claim 1, further comprising:
-   determining at least one error limit for at least one of said entities,
-   considering said at least one error limit during said determination of said matching measure.

4. Method according to claim 2, further comprising, allocating a position label to each of said entities in the string, and numbering same entities according to their relative position in accordance with the position label.

5. Method according to claim 2, further comprising:
-   determining a first distance between said two data entities of consecutively following data entities in said first data string,
-   determining a second distance of said two data entities dertermined in said second data string,
-   determine a difference between said first and second distances, and

- considering said difference during said determination of said matching measure.

6. Method according to claim 1, further comprising:
- storing said second string together with said similarity measure.

7. Method according to claim 1, further comprising:
- determining a threshold for said similarity measure, and
- outputting said second string, if said determined similarity measure at least equals said threshold.

8. Method according to claim 7, further comprising:
- repeating said determination of said similarity measure with a number of second strings, and
- determining said threshold in correspondence with a number of second strings to be outputted.

9. Method according to claim 1, further comprising:
- analyzing the first string for entities not present in the first string, and
- suppressing in the second string all said entities not present in said first string.

10. Method according to claim 9, further comprising:
- determining the number of entities that are present in the second string, but are not present in the first string, as a second similarity measure.

11. Method according to claim 10, further comprising:
- determining a section within said second string comprising at least the same number of entities that are simultaneously present in both strings.

12. Software tool comprising program code means stored on a computer readable medium for carrying out the method of anyone of claims 1 to 12 when said software tool is run on a computer or network device.

13. Computer program product comprising program code means stored on a computer readable medium for carrying out the method of anyone of claims 1 to 12 when said program product is run on a computer or network device.

14. Computer program product comprising program code, downloadable from a server for

carrying out the method of anyone of claims 1 to 12 when said program product is run on a computer or network device.

15. Computer data signal embodied in a carrier wave and representing a program that instructs a computer to perform the steps of the method of anyone of claims 1 to 12.

16. Electronic device for determining and outputting a similarity measure between two data strings each comprising data entities, comprising:
- a component for receiving a first string of entities and a second string of entities,
- a processing unit being connected to said receiving component, said processing unit being configured to determine at least one tuple of consecutively following data entities in said first data string, said processing unit being configured to determine the relative position of said at least one tuple of consecutively following data entities in said first data string, said processing unit being configured to determine at least one tuple of similar consecutively following data entities in said second data string, said processing unit being configured to determine the relative position of said at least one tuple of similar consecutively following data entities in said second data string, said processing unit being configured to determine a matching measure by comparing how far the relative positions of the at least one tuple of similar consecutively following data entities in said first data string matches with the relative position of said at least one tuple of similar consecutively following data entities in said second data string, and said processing unit being configured to output a similarity measure which corresponds to the matching measure of at least one comparison result, and
- an interface being connected to said for processing unit for outputting said similarity measure.

17. Electronic device according to claim 17, further comprising a storage connected to said processing unit for storing received strings and said determined similarity measures.